Combination of Gaussian Processes and Neural Network

Jiale Shi
Department of Chemical and Biomolecular Engineering
University of Notre Dame
Notre Dame, IN 46556
Jiale.Shi.33@nd.edu

Abstract

Gaussian processes (GPs) are generally nonparametric, which takes an infinite number of basis functions given by a probability distribution over functions. For a non-parametric model, the required number of parameters grows with the size of the training data; therefore, GPs have a very big challenge in the computational tractability. Inference in a Gaussian process has computational complexity of $\mathcal{O}(n^3)$. Even if using low rank approximations, the computational complexity is still $\mathcal{O}(n^2m)$, m is a user chosen parameter. A neural network (NN) is a parameterised model that can be tuned via gradient descent to approximate a labelled collection of data with high precision. The computational complexity of NN is much less than GPs. However, in the process of NN, the uncertainty would be dropped in the hidden layers. The uncertainty is very important because if the the uncertainty is missing, the high precision is meaningless. In addition, overfitting is a very common problem in NN. While GPs can track the uncertainty according to the prior. Therefore, combing the advantages of two worlds, GPs and NN together can be very useful to speed up computation and track uncertainty. Recently, there are some advancements in connecting GPs and NN.

1 Introduction

talk about the background of Neural Network and Gaussian Process

1.1 Style

2 Model

Show some models, like Deep Gaussian, Conditional Neural Process, Neural Process, Attentive Neural Process. and just show the 1D regression results.

An encoder h from input space into representation space that takes in pairs of context values (x_i, y_i) and produces a representatin $r_i = h((x_i, y_i))$ for each of the pairs. We parameterise h as a neural network.

An aggregator a that summarises the encoded inputs. Crucially, the aggregator reduces the runtime to $\mathcal{O}(n + m)$ where n and m are the number of context and target points respectively.

A conditional decoder g that takes as input the sampled global latent variable z as well as the new target locations x_T and outputs the predictions $\hat{y}_T$ for the corresponding values of $f(x_T) = y_T$.

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

3 Results

4 Related work

conditional neural processes [1], attentive neural processes [2].

5 Discussions

Please prepare

recently, deepmind published attentive neural processes paper beyond neural processes.

Acknowledgments

I would like to thank Prof. Nicholas Zabaras for giving wonderful lectures. I would like to thank TA Dr. Souvik Chakraborty for giving assistants. I would like to thank Navid Shervani-Tabar, Sina Malakpour, Govinda Anantha Padmanabha, Nick Geneva, Xian Gao and Elvis A. Eugene for insightful discussions.

References

[1] M. Garnelo, D. Rosenbaum, C. J. Maddison, T. Ramalho, D. Saxton, M. Shanahan, Y. W. Teh, D. J. Rezende, and S. Eslami. Conditional neural processes. *arXiv preprint arXiv:1807.01613*, 2018.

[2] H. Kim, A. Mnih, J. Schwarz, M. Garnelo, A. Eslami, D. Rosenbaum, O. Vinyals, and Y. W. Teh. Attentive neural processes. *arXiv preprint arXiv:1901.05761*, 2019.